

**AMENDMENTS TO THE DRAWINGS:**

The attached sheet of drawings includes changes to Figure 5. This sheet replaces the original sheet including Figure 5. In Figure 5, previously omitted elements 20, 22 and 26 have been added.

Attachment:	Replacement Sheet
	Annotated Sheet Showing Changes

### **REMARKS**

In amended Figure 5, the previously omitted elements numerals 20, 22 and 26 have been added.

Claims 1-20 are pending in this application. Claims 1 and 11 have been amended.

Claims 3 and 4 were rejected under 35 US §112, second paragraph. Claims 3 and 4 have not been amended since Applicant believes the limitation “structured document” in each claim is correct in that it refers back to the language in the preamble of independent Claim 1: a method for creating a structured document. Claim 3 is directed to the method of Claim 1, wherein the set of tag suggestions are generated during creation of the structured document. Claim 4 is directed to the method of Claim 1, wherein the set of tag suggestions are generated prior to creation of the structured document.

Claims 1-8, 11-20 were rejected under 35 USC §103(a) as being unpatentable over Cowan et al. (“RITA – An Editor and User Interface for Manipulating Structured Documents”; Pub Data 1991) in view of Lindbald et al. (Publication No. US 2004/0103091 A1; Filing Date June 13, 2003) and further in view of Brockway et al. (Publication No. US 2004/0249795 A1; Filing Date June 5, 2003).

Claims 9 and 10 were rejected under 35 USC §103(a) as being unpatentable over Cowan et al. (“RITA – An Editor and User Interface for Manipulating Structured Documents”; Pub Data 1991) in view of Lindbald et al. (Publication No. US 2004/0103091 A1; Filing Date June 13, 2003) and further in view of Brockway et al. (Publication No. US 2004/0249795 A1; Filing Date June 5, 2003) and further in view of Kotsakis (“Structured Information Retrieval in XML Documents”; Publication Date 2002).

Claim 1 is directed to method for converting a generic document, wherein a generic document comprises a document in a particular format type, into a structured document, wherein a structured document includes a plurality of content elements wrapped in pairs of hierarchically nested tags, comprising: parsing the document of the particular format type containing content into a plurality of content elements; and for a selected content element, suggesting an optimal tag according to a tag suggestion procedure; wherein the tag suggestion procedure comprises: providing sample data in the form of structured sample documents; analyzing patterns in the

sample data to derive a set of tag suggestions; deriving a set of candidate tags from the set of tag suggestions for the selected content element; and evaluating the set of candidate tags according to tag suggestion criteria to determine an optimal tag for the selected content element. Claim 1 claims a method for converting existing document in “generic” formats into structured documents. The number of documents available/generated in XML format remains fairly low as compared to documents in other formats in part because converting documents from other formats into XML is often difficult and time-consuming. The method of Claim 1 provides a method for converting existing documents into structured documents.

RITA teaches a method of authoring new structured documents. RITA teaches creating structured documents using pre-defined document classes (each document class has a particular document structure, which is displayed to the user in a user interface). “RITA was designed for users with a wide range of expertise in document processing and has the primary goal that members of this broad class can create and edit documents which are structurally correct as they are initially entered into the computer.” See second paragraph of the Introduction of RITA. Apparently RITA can be used only to create new structured documents or to edit old structured documents (e.g., structured documents that have been originally created using RITA). In contrast, Applicant’s method of Claim 1 provides a method of converting existing generic documents into structured documents. RITA does not teach or suggest “parsing the document of the particular format type containing content into a plurality of content elements”.

RITA accepts input from two sources, a document class database and the user. See the third paragraph of the System Overview of RITA:

Documents such as papers, memoranda, and letters have different syntactic structures and are regarded as structured objects, with each type of document defining a document class. Thus a particular document is an instance of a specific document class. The database of document class descriptions provides information on the structure to which a document of a given class must conform, as well as how concrete representations of such document must be generated. These class descriptions or document types are normally created by a document administrator or expert user.

Figure 3 of RITA shows how the user can create a document of a particular document class. The structure for the defined document class is shown on the left of the interface. The user creates the text of the document by typing in the window on the right. RITA does not teach or suggest

“parsing the document of the particular format type containing content into a plurality of content elements”.

RITA suggests tags to users through the user interface (Fig. 3). These tag suggestions are based on the user selecting a particular document class. The tag suggestions for each document class are created by a document administrator or expert user and displayed in a window for the user to select. The user must guess the appropriate tag suggestion for each part of the document he is creating. In contrast, in Applicant's method, the tag suggestion procedure comprises: providing sample data in the form of structured sample documents; analyzing patterns in the sample data to derive a set of tag suggestions; deriving a set of candidate tags from the set of tag suggestions for the selected content element; and evaluating the set of candidate tags according to tag suggestion criteria to determine an optimal tag for the selected content element.

Lindblad describes a structural textual classification system for classifying elements in an XML database. The Examiner cited paragraph 56 of Lindblad:

In the preceding examples, the data sets are text documents. Much research and many systems are know for characterizing test [sic] document, and analyzing the resulting qualities. For example, a search engine characterizes text document by their content words, relative placement and occurrence of words and possibly their storage location and an analysis process selects text documents by whether or not they meet some specified search query constraint. In many cases, where the data sets are XML documents, they are often characterized as if they were just text documents. This does not yield desirable results in many cases.

While Lindblad recognizes that many text documents have been analyzed for their resulting qualities, Lindblad suggests that characterizing XML (i.e., structured) documents in the same way “does not yield desirable results in many cases”. Thus it would appear that Lindblad teaches away from “providing sample data in the form of structured sample documents” as part of a tag suggestion process: “wherein the tag suggestion procedure comprises: providing sample data in the form of structured sample documents; analyzing patterns in the sample data to derive a set of tag suggestions; deriving a set of candidate tags from the set of tag suggestions for the selected content element; and evaluating the set of candidate tags according to tag suggestion criteria to determine an optimal tag for the selected content element.”

Brockway teaches comparing a test document tree to a model document tree to determine if the test document's semantics match the model document's. Brockway describes a system for

searching for information in a distributed data processing system, including providing a semantics-based search index by storing in the semantics-based search index search key-words from documents according to semantics from selected document structure templates, the document structure templates selected in dependence upon the structures of the documents and upon model document structures in the document structure templates. See the Abstract. Brockway is concerned with the problem of returning better search results. Brockway wants to improve search results so the search results take into account semantics. In paragraph 84 of Brockway, Brockway describes scanning through a tree structure for the [current] document and through a tree structure of a model document structure, comparing the two trees, incrementing an integer score for all points of positive comparison, and returning a score value. In paragraph 86 of Brockway, Brockway states that if the score exceeds the threshold, runHeuristic concludes that the semantics of its document structure template are good for the current document and sets a reference to its document structure template in the indexing search engine. In contrast, Applicant's method includes, in part, "providing sample data in the form of structured sample documents; analyzing patterns in the sample data to derive a set of tag suggestions; deriving a set of candidate tags from the set of tag suggestions for the selected content element."

None of RITA, Lindblad or Brockway, whether taken alone or in combination, teaches or suggests "for a selected content element, suggesting an optimal tag according to a tag suggestion procedure." In RITA, the user must select the tag from the list displayed on the left side of the interface. Lindblad is concerned with classifying elements in an XML database. Brockway compares document trees for the purpose of evaluating a document's semantics, making no tag suggestion in response to a selected content element.

Claim 11 is directed to a method for predicting structure and content during authoring of a structured document (i.e., without providing an initial document). Claim 11 claims a method for authoring of a structured document, wherein a structured document comprises a plurality of content elements wrapped in pairs of tags, comprising: generating content elements wrapped in pairs of tags; and for a selected tag, suggesting an optimal content fragment according to a content suggestion procedure; wherein the content suggestion procedure comprises: providing a sample structured document; deriving a set of content fragments from the sample structured

document; evaluating the set of content fragments according to a content fragment suggestion criteria to determine an optimal content fragment suggestion for the tag, wherein the optimal content fragment suggestion is the most probable content fragment for the selected tag.

As noted above, RITA teaches a method of authoring new structured documents (and editing existing structured documents. RITA teaches creating structured documents using pre-defined document classes (each document class has a particular document structure, which is displayed to the user in a user interface). Nothing in RITA teaches or suggests “for a selected tag, suggesting an optimal content fragment according to a content suggestion procedure. None of Lindblad or Brockway teaches or suggests “suggesting an optimal content fragment” in response to a selected tag element.

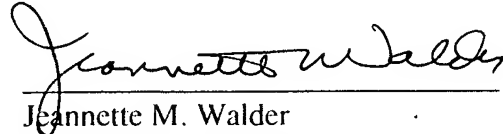
Claims 1 and 11 are believed to be patentable over the references cited. Since Claims 2-10, and 20 depend from Claim 1 and Claims 12-19 depend from Claim 11, they are also believed to be patentable.

No additional fee is believed to be required for this amendment; however, the undersigned Xerox Corporation attorney hereby authorizes the charging of any necessary fees, other than the issue fee, to Xerox Corporation Deposit Account No. 24-0025.

Application No.: 10/607,667

Reconsideration of this application and allowance thereof are earnestly solicited. In the event the Examiner considers a personal contact advantageous to the disposition of this case, the Examiner is requested to call the undersigned Attorney for Applicant, Jeannette Walder.

Respectfully submitted,

A handwritten signature in cursive script, reading "Jeannette M. Walder", written over a horizontal line.

Jeannette M. Walder  
Attorney for Applicant  
Registration No. 30,698  
Telephone: 714-565-1700

Xerox Corporation  
Santa Ana, California  
Date: June 30, 2006

Attachments

4/5

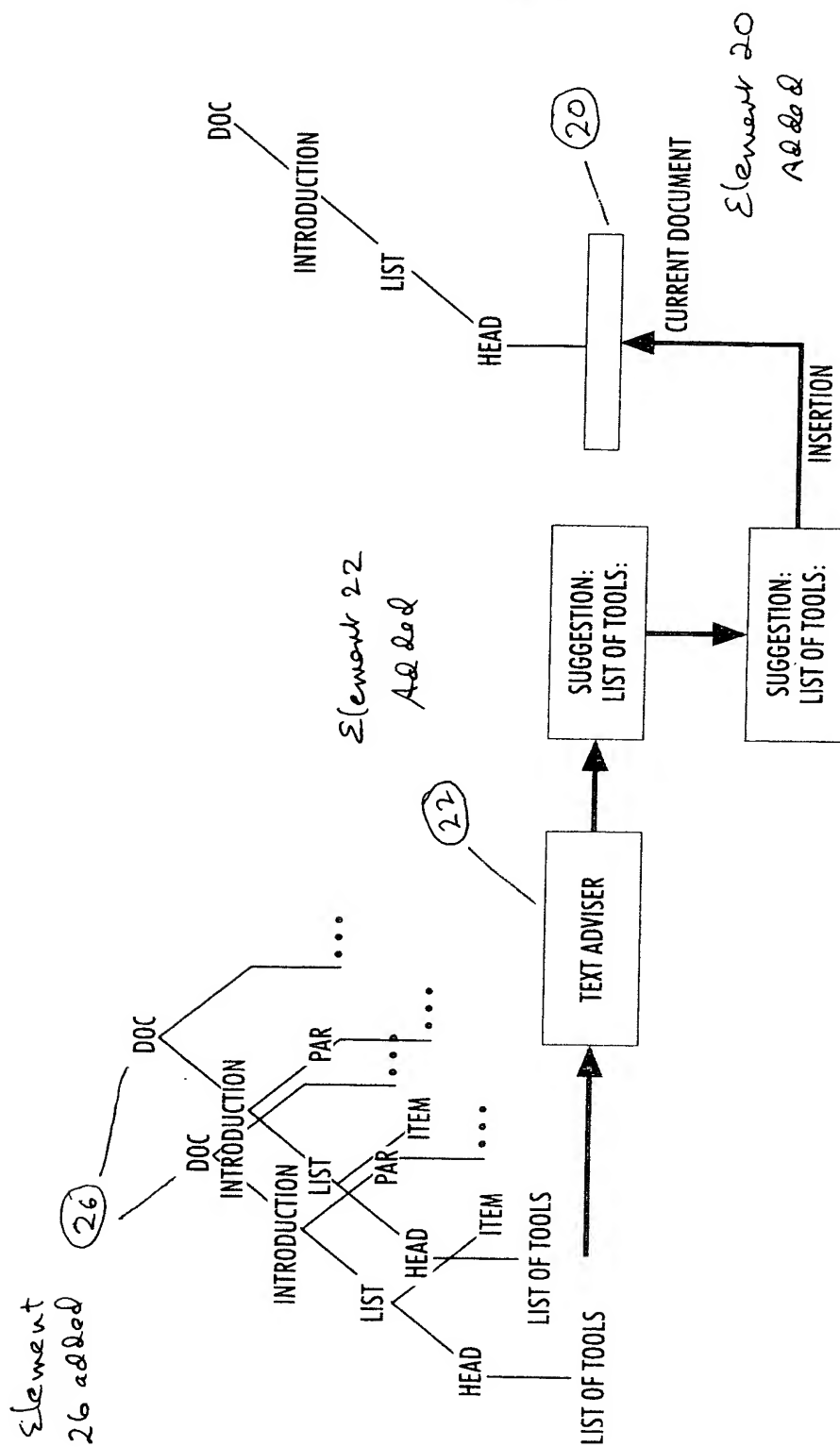


FIG. 5